

Integrated proteome and phosphoproteome analyses of peripheral blood mononuclear cells in primary Sjögren syndrome patients

Shaoying Huang¹, Fengping Zheng¹, Lixiong Liu², Shuhui Meng¹, Wanxia Cai¹, Cantong Zhang¹, Weier Dai⁴, Dongzhou Liu², Xiaoping Hong², Donge Tang¹, Yong Dai^{1,3}

¹Department of Clinical Medical Research Center, Guangdong Provincial Engineering Research Center of Autoimmune Disease Precision Medicine, Shenzhen People's Hospital, The First Affiliated Hospital Southern University of Science and Technology, The Second Clinical Medical College of Jinan University, Shenzhen 518020, Guangdong, China

²Department of Rheumatology and Immunology, The Second Clinical Medical College of Jinan University, The First Affiliated Hospital of Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen 518020, Guangdong, China

³Guangxi Key Laboratory of Metabolic Disease Research, Nephrology Department of Guilin, Guilin 541002, China

⁴College of Natural Science, University of Texas at Austin, Austin, TX 78712, USA

Correspondence to: Yong Dai, Donge Tang, Xiaoping Hong; email: daiyong22@aliyun.com, <https://orcid.org/0000-0002-6840-9158>; dongeg66@126.com, <https://orcid.org/0000-0001-7011-196X>; hongxiaoping123@tom.com, <https://orcid.org/0000-0002-2062-8394>

Keywords: primary Sjögren syndrome, peripheral blood mononuclear cells, phosphoproteome, biomarker

Received: June 18, 2020

Accepted: October 27, 2020

Published: December 3, 2020

Copyright: © 2020 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Primary Sjögren syndrome (pSS) is a common autoimmune disease. Here, we performed the first proteome and phosphoproteome analyses of peripheral blood mononuclear cells in pSS patients to obtain a comprehensive profile and identify the potential crucial proteins and pathways for the screening and evaluation of pSS patients. Peripheral blood mononuclear cells from 8 pSS-confirmed patients (American-European Consensus Group Criteria, 2002) and 10 normal controls were selected. Label-free quantitative proteomics was utilized to obtain quantitative information. In total, 787 proteins were identified as differentially expressed proteins, and 175 phosphosites on 123 proteins were identified as differentially phosphorylated proteins. We performed functional enrichment analyses with these proteins and phosphoproteins based on public database. Furthermore, protein-protein interaction network analyses were performed by using multiple algorithms. Using module and hub protein analyses, we identified 16 modules for the proteins, 2 clusters for the phosphoproteins and selected the top 10 hub proteins. Finally, we identified 22 motifs using motif analysis of the phosphosites and found 17 newly identified motifs, while 6 motifs were experimentally verified for known protein kinases. The findings distinguished pSS patients from normal controls at the peripheral blood mononuclear cells level and revealed potential candidates for use in pSS diagnosis.

INTRODUCTION

Primary Sjögren syndrome (pSS) is a chronic autoimmune disease with lymphocytic infiltration and epithelial cell destruction but no association with other autoimmune diseases (e.g., scleroderma, systemic lupus erythematosus or rheumatoid arthritis) [1]. Clinically, it

often invades the salivary gland and lacrimal gland, where it manifests as xerostomia of the mouth and eyes [1], which might severely decrease patients' quality of life. More seriously, the function of exocrine glands, such as those in the respiratory system, digestive system, skin and vagina, are also often damaged, and lesions outside the glands can also be seen [1–3]. Patients with visceral

damage mostly achieve remission after proper treatment, but they can also relapse after treatment withdrawal [4]. Furthermore, patients with progressive pulmonary fibrosis, renal tubular acidosis, central neuropathy, pulmonary hypertension, lymphoma in visceral damage, and acute pancreatitis have a poor prognosis [4]. With an overall prevalence of 0.1-0.6% in the adult population [5], nearly 0.1% in Europe [6] and 0.33-0.77% in China [7], no disease-modifying drug is currently available for the disease, and the current treatment for pSS focuses only on relieving symptoms, which is unsatisfactory [8].

Early diagnosis ensures timely management of both the clinical presentation and complications. The revised American-European classification criteria (AECG) requires a positive minor gland biopsy sample, autoantibodies, and symptoms with oral and ocular dryness for the diagnosis of SS [9]. Typical symptoms, however, may occur several months to years before the ultimate diagnosis. Therefore, serological markers are desperately needed, such as anti-Ro/SSA [10] and anti-La/SSB [11], for a provisional diagnosis. However, the autoantibody markers currently available cannot satisfy the need for specificity and/or sensitivity for a confirmatory diagnosis [12]. Additionally, labial salivary gland biopsy has become the most widely accepted test to confirm pSS. However, its limitations are obvious; for example, a biopsy is an invasive method, and there is no observable lymphocytic infiltration in some of the confirmed subjects [13]. Consequently, there is an unmet need for noninvasive clinical tests to identify pSS.

The emerging thoughts on precision medicine underscores the importance of quantitative portrayals of molecular features. Notably, peripheral blood mononuclear cells (PBMCs) consist of immune cells, including natural killer (NK) cells from the innate immune system or lymphocytes from the adaptive immune system or both systems (monocytes and dendritic cells). Thus, PBMCs have attracted increasing attention as intermediaries in autoimmune disease and proteomic studies, which may facilitate the investigation of etiopathogenesis factors or predictors of pSS. Compared with biopsy, PBMCs can be conveniently and safely obtained by isolation from peripheral blood. In addition, genomic data provide a considerable amount of information to decipher regarding molecular machinery. However, the expression and functional details of genes are often best reflected by the presence (or absence) of proteins acting as executors of physiological functions. Protein phosphorylation is a reversible posttranslational modification that is also one of the major mechanisms of signal integration and is essential for the regulation and maintenance of most biological processes in eukaryotes [14]. Phosphorylation is employed by cells to transiently alter protein properties such as the activities of enzymes

and interactions with other proteins [15]. Thus, we took the initiative to perform the first analyses of the PBMC proteome and phosphoproteome of pSS.

In this study, differentially expressed proteins (DEPs) and differentially phosphorylated proteins (DPPs) were identified by label-free quantitative shotgun proteomic analyses, and phosphopeptide enrichment was determined using immobilized metal affinity chromatography (IMAC) using PBMCs. The elaborate bioinformatics strategies, including Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway analysis, motif analysis, and protein-protein interaction (PPI) network analysis were utilized to analyze the data and reveal pSS-related modulation and pSS signatures both in the proteome and phosphoproteome, thereby contributing to an initial PBMC-based map of pSS. The novel DEPs and DPPs identified in the PBMCs of pSS may shed light on biomarkers useful for disease screening and the underlying molecular mechanisms for use in eventually tackling pSS.

RESULTS AND DISCUSSION

Qualitative and quantitative proteomic profiling

The schematic flow is shown in Figure 1A. Based on a shotgun proteomics approach, 2517 proteins were identified in the study based on 12692 unique peptides and a maximum false positive rate (FDR) < 1%, among which 1111 proteins were at a quantifiable level in the pSS or NC-pSS groups (Supplementary Table 1). The combined results were used for the following analyses. Considering the criterion of ratio change > 2-fold or < 0.5-fold for proteins to be considered significantly differentially expressed, 787 proteins were identified as differentially expressed proteins (DEPs). Among these proteins, 661 were up-regulated, and 126 were down-regulated in the pSS group (Figure 1B). The top 20 proteins up-regulated in pSS are listed in Table 1. These 20 proteins showed the most significant differences in the quantitative comparison of healthy volunteers and pSS patients at the PBMC level and, therefore, appeared to be the potential discriminatory biomarkers for pSS.

Additionally, there were some preliminary proteins that were not classified as DEPs under the criterion of ratio change > 2-fold, such as HLA-DRA (ratio: 1.751) and HLA-DRB1 (ratio: 1.528), but were considered as previous potential biomarkers for Sjögren syndrome. Studies have shown that the major histocompatibility complex (MHC) gene frequency in Sjögren syndrome patients increased, and the positive rate of HLA-B8, DR3 and DRw52 gene expression was significantly higher than that in the normal subjects [4]. We

identified HLA-B (ratio: 10.429, up-regulated) and HLA-A (ratio: 6.38, up-regulated) in the PBMC proteome, which corroborated the high frequency of the MHC genes found in pSS patients. HLA-A is associated with increased susceptibility to multiple autoimmune diseases, such as multiple sclerosis (MS) [16] and Behcet disease (BD) [17]. HLA-A has not been proven to be associated with pSS and is reported in the PBMC proteome level for the first time.

GO classification of the DEPs

The GO classifications of the proteins in the pSS proteomes are shown in Figure 1C. The proteins presented a similar distribution of DEPs in pSS based on the GO classification. In the cellular component classification, most of the proteins were in the cell, organelle, extracellular region, membrane, and macromolecular complex categories. In the molecular function classification, the majority of the proteins were

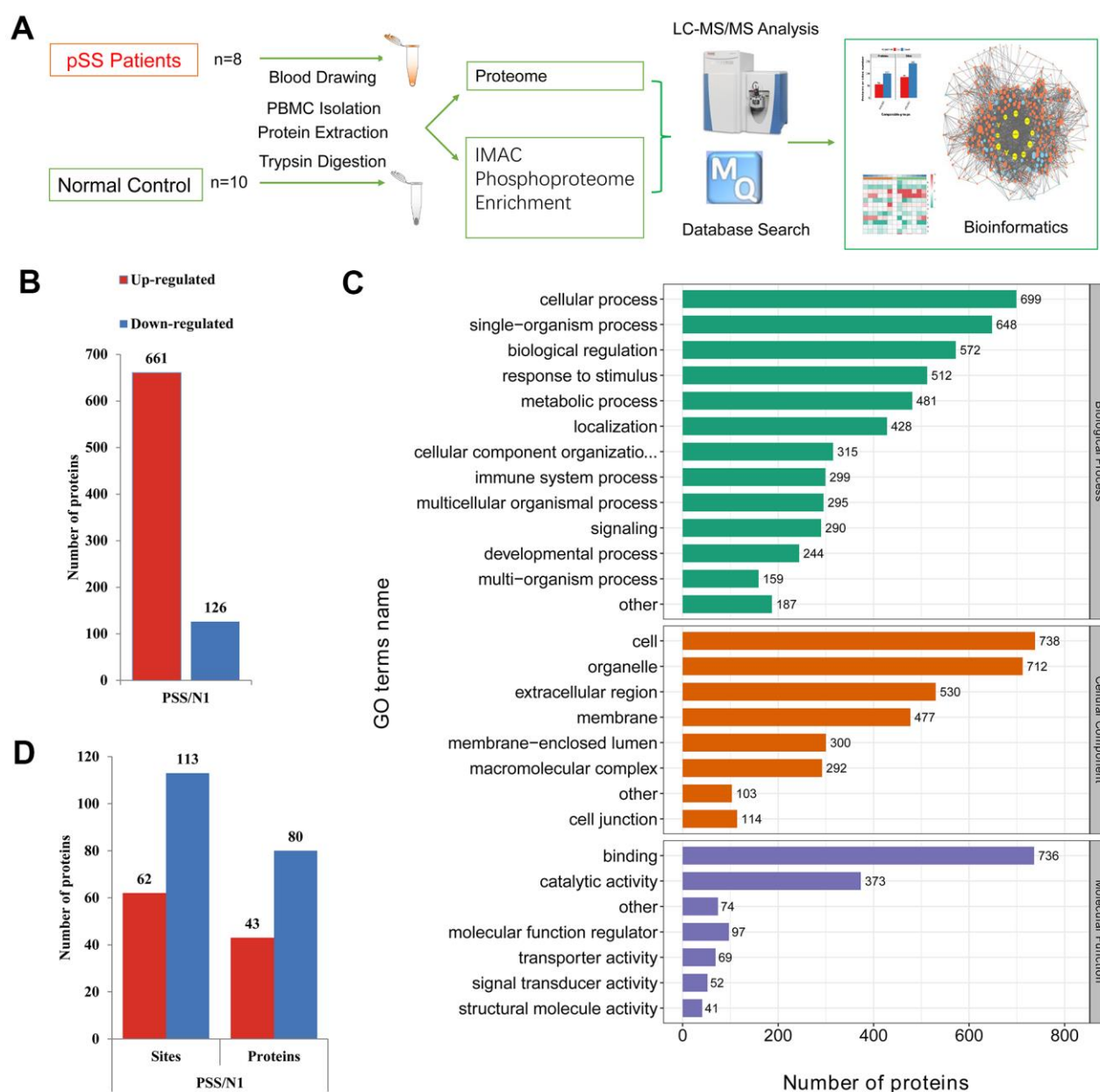


Figure 1. (A) The schematic flow to study the proteome and phosphoproteome of peripheral blood mononuclear cells (PBMCs) in primary Sjögren syndrome (pSS) patients. (B) Histogram of quantity distribution of differentially expressed proteins (DEPs). (C) Gene Ontology (GO) classifications of the DEPs of PBMCs in pSS patients based on biological process, cellular component and molecular function. (D) Histogram of quantity distribution of differentially expressed phosphorylated proteins (DPPs). IMAC: immobilized metal affinity chromatography. LC-MS: liquid chromatography–mass spectrometry.

Table 1. The summary of top 20 upregulated differentially expressed proteins.

Protein accession	Gene name	PSS/N1 Ratio	Regulated Type
Q13813	SPTAN1	61.5	Up
O43150	ASAP2	43.444	Up
Q9UBW5	BIN2	23.691	Up
Q9Y2A7	NCKAP1	20.978	Up
O43149	ZZEF1	20.277	Up
P42025	ACTR1B	18.802	Up
P09960	LTA4H	17.519	Up
P48637	GSS	16.094	Up
P48059	LIMS1	14.748	Up
Q9C0I1	MTMR12	14.625	Up
P05556	ITGB1	14.385	Up
Q9Y3P9	RABGAP1	14.152	Up
O75791	GRAP2	13.599	Up
Q05193	DNM1	12.699	Up
Q4KMQ2	ANO6	12.333	Up
Q9C0C9	UBE2O	12.158	Up
Q96KP4	CNDP2	11.821	Up
P80511	S100A12	11.739	Up
O95373	IPO7	11.739	Up
P52788	SMS	11.422	Up

in the binding, catalytic activity, and molecular function regulator categories. In the biological process classification, most proteins were observed to be involved in cellular processes, single-organism processes, and biological regulation. Notably, the immune system process category contained 299 DEPs (6%). Among them, LTA4H (ratio: 17.519, up-regulated), S100A12 (ratio: 11.739, up-regulated), CD36 (ratio: 6.194, up-regulated), and PRKCB (ratio: 4.682, up-regulated) being involved in inflammatory processes and immune response caught our attention. LTA4H catalyzes the final step in the biosynthesis of the proinflammatory mediator leukotriene B4 [18]. S100A12 plays a prominent role in the regulation of inflammatory processes and immune response. Its proinflammatory activity involves recruitment of leukocytes, promotion of cytokine and chemokine production, and regulation of leukocyte adhesion and migration [19]. CD36 is a multifunctional glycoprotein

that acts as a receptor for a broad range of ligands. Ligands can be of proteinaceous nature like TLR4, which promotes inflammation in monocytes/macrophages [20]. PRKCB is a protein kinase involved in regulation of the B-cell receptor (BCR) signalosome, and plays a key role in B-cell activation [21]. These proteins were specifically up-regulated, suggesting a role of immunology and inflammation in the PBMC proteome of pSS.

GO, KEGG and protein domain functional enrichment analyses and clustering analysis of the DEPs

The results of the GO enrichment analysis are shown in Figure 2A and 2B. Regarding the enrichment of proteins based on biological process, negative regulation of hydrolase activity and regulation of cytoskeleton organization were significantly enriched. Regarding the enrichment of the proteins that

were up-regulated in the pSS group based on biological processes, the regulation of cytoskeleton organization and actin cytoskeleton organization were enriched. Regarding the enrichment of the proteins that were down-regulated in the pSS group based on biological process, the categories of protein activation cascade, immunoglobulin mediated immune response, B-cell-mediated immunity, regulation of protein maturation, regulation of humoral immune response, regulation of protein activation cascade, acute inflammatory response, regulation of inflammatory response, activation of immune response, and positive regulation of immune response were enriched.

Regarding the enrichment of proteins that were up-regulated in the pSS group based on molecular function, categories such as protein complex binding, actin binding and cell adhesion molecule binding were enriched. In the molecular function classification, the peptidase inhibitor activity, serine-type endopeptidase inhibitor activity, endopeptidase regulator activity, and enzyme inhibitor activity categories had a significant abundance in pSS down-regulated proteins.

In the cellular component classification, the cytoplasmic vesicle lumen, vesicle lumen, and secretory granule lumen categories had a significant abundance of pSS proteins. Regarding the enrichment of the proteins that were up-regulated in the pSS group based on the cellular component classification, categories such as ficolin-1-rich granules were enriched. According to the enrichment of the proteins that were down-regulated in the pSS group based on the cellular component classification, categories such as extracellular space, blood microparticle, and extracellular region were enriched.

For KEGG enrichment of DEPs, we identified 16 pathways from up-regulated DEPs (Figure 2C) and 8 pathways from down-regulated DEPs (Figure 2D). The pathway of complement and coagulation cascades (hsa04610) was substantially down-regulated as an immunological pathway among the identified pathways. All of the significantly enriched pathways, namely, cGMP-PKG signaling pathway (hsa04022), arrhythmogenic right ventricular cardiomyopathy (ARVC) (hsa05412), hypertrophic cardiomyopathy (HCM) (hsa05410), and vascular smooth muscle contraction (hsa04270) pathways, seemed to be related to the cardiovascular system, and the cardiovascular risk increased in pSS, although the underlying mechanisms remain unclear [22]. It is not clear whether these pathways play direct roles in the mechanisms of pSS.

The protein domain refers to specific components that appear repeatedly in different protein molecules, which have similar sequences, structures and functions and are the units conserved during protein evolution. The length of the domain is usually between 25 and 500 amino acids in length. We identified the enrichment of the significant protein domains in up-regulated DEPs, such as thioredoxin domain, thioredoxin-like fold, EF-hand domain, and pleckstrin homology domain.

Finally, we divided all of the identified proteins into four quantiles (Q1-Q4) according to fold change: Q1 ($0 < \text{ratio} < 0.5$), Q2 ($0.5 < \text{ratio} < 0.67$), Q3 ($1.5 < \text{ratio} < 2.0$), and Q4 ($\text{ratio} > 2.0$) and performed functional enrichment clustering analyses (Supplementary Figures 1–5).

PPI network of the DEPs

Genes causing related diseases tend to cluster with one another in an interaction network of proteins or functions [23]. Therefore, the PPI network identifies and characterizes related protein complexes, which are crucial for understanding the molecular events involved [24]. We determined the PPI relationships using STRING [25]. The interactions in the STRING database were based on coexpression, text mining, databases, experiments, cooccurrence, neighborhoods, and gene fusion [26]. To show the interaction between the proteins clearly, we selected the top 50 proteins with the closest interactions and generated a network with a total number of 395 associated proteins. We selected the top 10 proteins as hub proteins based on degree (Table 2), which often play important roles in network stability because of their high degree of connections/interactions. We found that 8 out of the top 10 are up-regulated and involved in immune system processes.

MAPK1 (ratio: 2.676, up-regulated) had a high degree of 69 of interrelatedness, suggesting its potential for having an important role in pSS. MAPK1 is a member of a superfamily of mitogen-activated protein kinases (MAPKs), which control a variety of functions in eukaryotic cells [27, 28]. Altered MAPK signaling has been associated with autoimmune diseases such as lupus [29]. However, the association between MAPK signaling and pSS has been poorly characterized.

TGFB1 (ratio: 3.819, up-regulated) is a multifunctional protein that regulates the growth and differentiation of various cell types and is involved in various processes, such as normal development and immune function [30]. It was involved in inflammatory bowel disease (IBD) with impaired T-cell response to stimulation and decreased T-cell subsets [31]. Interestingly, peripheral blood T cell reduction is one of the most prominent

characteristics of immune abnormalities in pSS patients. However, the relationship between TGFB1 and pSS has not been well elucidated yet.

Module analyses

To conduct PPI topological analysis, we combined the DEPs that are known biomarkers of pSS and applied them to the molecular complex detection

(MCODE) plugin in Cytoscape [31]. MCODE is one of the most commonly used PPI clustering algorithms and can isolate dense regions and predict protein complexes via a PPI subnetwork based on connectivity data [25, 32]. MCODE yielded 16 modules, and the 3 most significant DEP modules were extracted. Module 1 (MCODE score =26.000) was constructed with 26 nodes and 325 edges (Figure 3A);

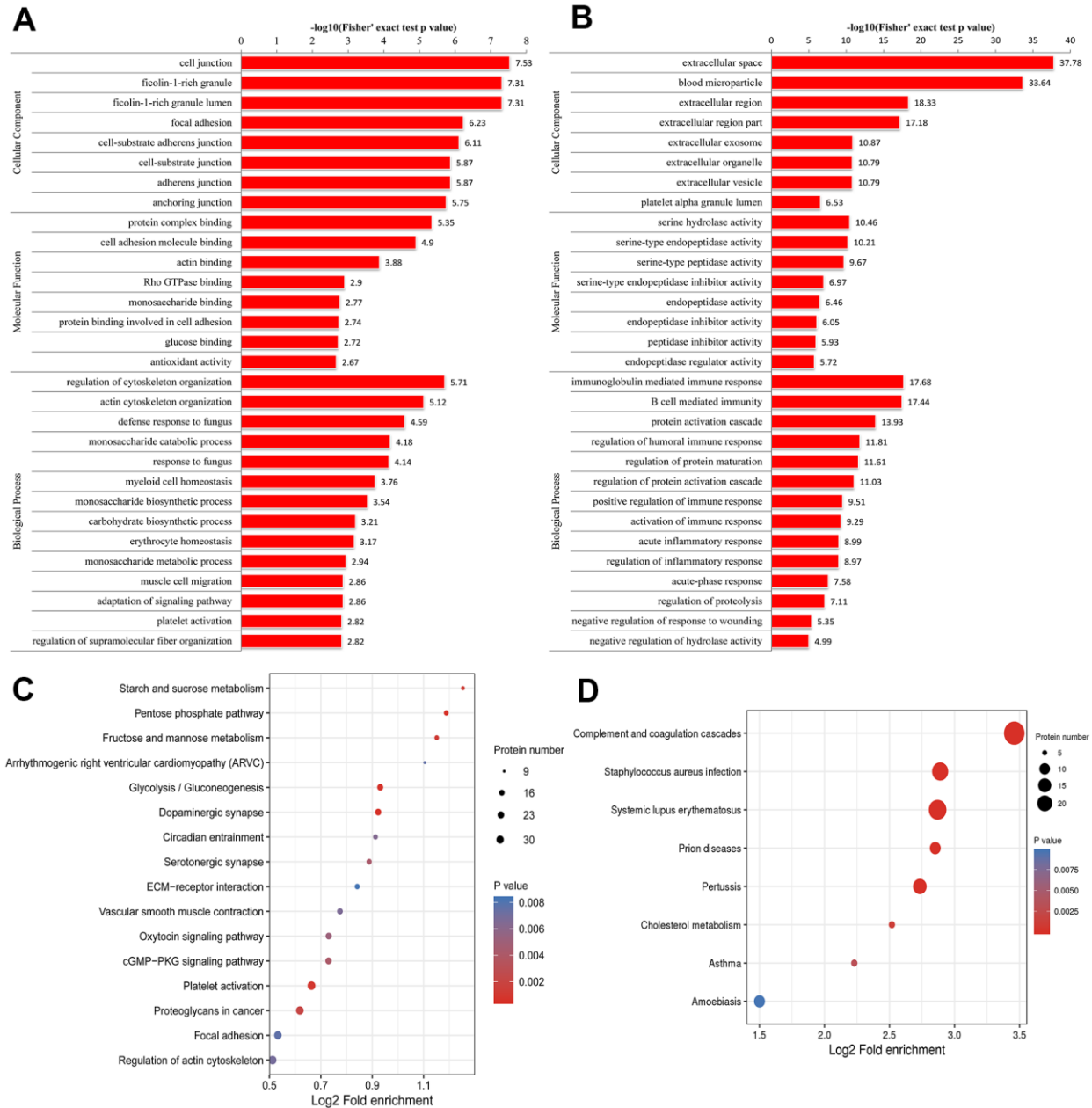


Figure 2. Gene Ontology (GO) enrichment analysis of peripheral blood mononuclear cells (PBMCs) in primary Sjögren syndrome (pSS) patients based on biological process, cellular component and molecular function. (A) GO enrichment analysis of up-regulated differentially expressed proteins (DEPs). (B) GO enrichment analysis of down-regulated DEPs. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis of up-regulated differentially expressed proteins (DEPs). (D) KEGG functional enrichment analysis of down-regulated DEPs.

Table 2. Top 10 hub proteins in PPI network based on degree.

Gene name	Degree	Betweenness	Centroid	Closeness	Bridging
UBA52	75	23277.55	-22	0.001153	32.12191
MAPK1	69	14795.83	9	0.001176	28.44867
HSPA8	56	14824.35	-14	0.001145	56.68246
ACTB	55	11072.69	-23	0.001142	40.24189
ACLY	52	14155.31	-159	9.90E-04	35.44649
APP	49	7938.799	-44	0.001119	36.50353
ACTN1	49	5651.568	-20	0.001144	38.92992
ACTN4	47	5555.22	-22	0.00114	43.87646
VWF	46	3615.55	-104	0.001048	35.02035
TGFB1	41	8255.129	-19	0.001151	58.37043

Degree, betweenness, centroid, closeness and bridging are algorithms provided by plugin Centiscape in software Cytoscape.

module 2 (MCODE score =17.758) was constructed with 34 nodes and 293 edges (Figure 3B); and module 3 (MCODE score =12.261) was constructed with 24 nodes and 141 edges (Figure 3C).

Proteome-wide analyses of the phosphorylated PBMCs in the pSS patients

The global phosphoproteomic profiles from the PBMCs of the pSS patients had not been previously reported. In this study, we performed quantitative phosphoproteomic analyses of the PBMCs from the pSS patients.

Identification of phosphopeptides

In total, the experiment revealed the identification of 4349 peptides. Of all the acquired peptides, 2148 phosphopeptides were obtained. To maintain high localization confidence, the data were filtered with the combined cutoff values for the identification of the localization probability > .75. In total, 2449 unique phosphosites on 1239 proteins were identified, among which the quantitative information of 175 sites on 123 proteins was obtained. For all the identified phosphosites, 84.95% were phosphorylated at a serine residue, 13.58% at a threonine residue and 1.47% at a tyrosine residue. The serine and threonine distribution of phosphosites in our study was consistent with that of the results from human phosphoproteomes determined *in vivo*, while our tyrosine phosphorylation proportion was slightly higher [15].

Differential phosphorylation may be the effect of changes in protein expression or changes of a phosphosite. To discern the cause of the differential findings, we normalized the site ratio with the protein ratio to determine the relative change in phosphory-

lation site stoichiometry based on the quantified proteomic data obtained, which facilitated the discernment of the changes from the modification itself and those in the pathways that modulate pSS [32]. The results were used for the following analyses. For the quantitative analysis, 62 phosphosites on 43 phosphoproteins were determined to be up-regulated while 113 on 80 phosphoproteins were down-regulated, considering the ratio change greater than 2-fold indicated upregulation while that < 2-fold indicated down-regulation, which suggested molecular diversity in the PBMCs of the pSS group and would be gathered to conduct the following bioinformatics analysis (Figure 1D). The down-regulated phosphoproteins were obviously more common than the up-regulated phosphoproteins, in contrast to the ratio observed in the proteome assessed in the current study. These identified phosphoproteins might be crucial and valuable factors in the mechanism of pSS.

GO and subcellular localization classification of the phosphoproteins

To investigate the phosphoproteome in pSS, we performed GO classification of the phosphoproteins. The biological process and cellular component categorization of the differentially regulated phosphoproteins revealed a similar enrichment pattern as that for the pSS proteome. The three largest classes of biological processes were cellular processes (12%), single-organism processes (12%) and biological regulation (10%). Cell (21%), organelle (19%), and membrane (17%) were the three major categories in the cellular component classification. Based on the molecular function classification, the three largest categories included binding (49%), catalytic activity (21%) and molecular function regulator (11%). Shown by the

analysis of subcellular localization, the 439 quantifiable phosphoproteins were distributed in the cytoplasm (37%), nucleus (27%) and plasma membrane (11%).

GO and KEGG functional enrichment analyses of the phosphoproteins

A GO enrichment analysis was performed to better understand the functions of the identified phosphoproteins. According to biological process enrichment classification, regulation of ion transmembrane

transport, protein complex assembly, secretion, and homotypic cell-cell adhesion were greatly enriched in up-regulated phosphoproteins. In contrast, platelet activation, blood coagulation, and cell morphogenesis involved in differentiation were greatly enriched in down-regulated phosphoproteins. In agreement with this observation, the molecular function enrichment analysis showed significant enrichment of binding categories such as kinase binding, identical protein binding, enzyme binding, integrin binding, and mitogen-

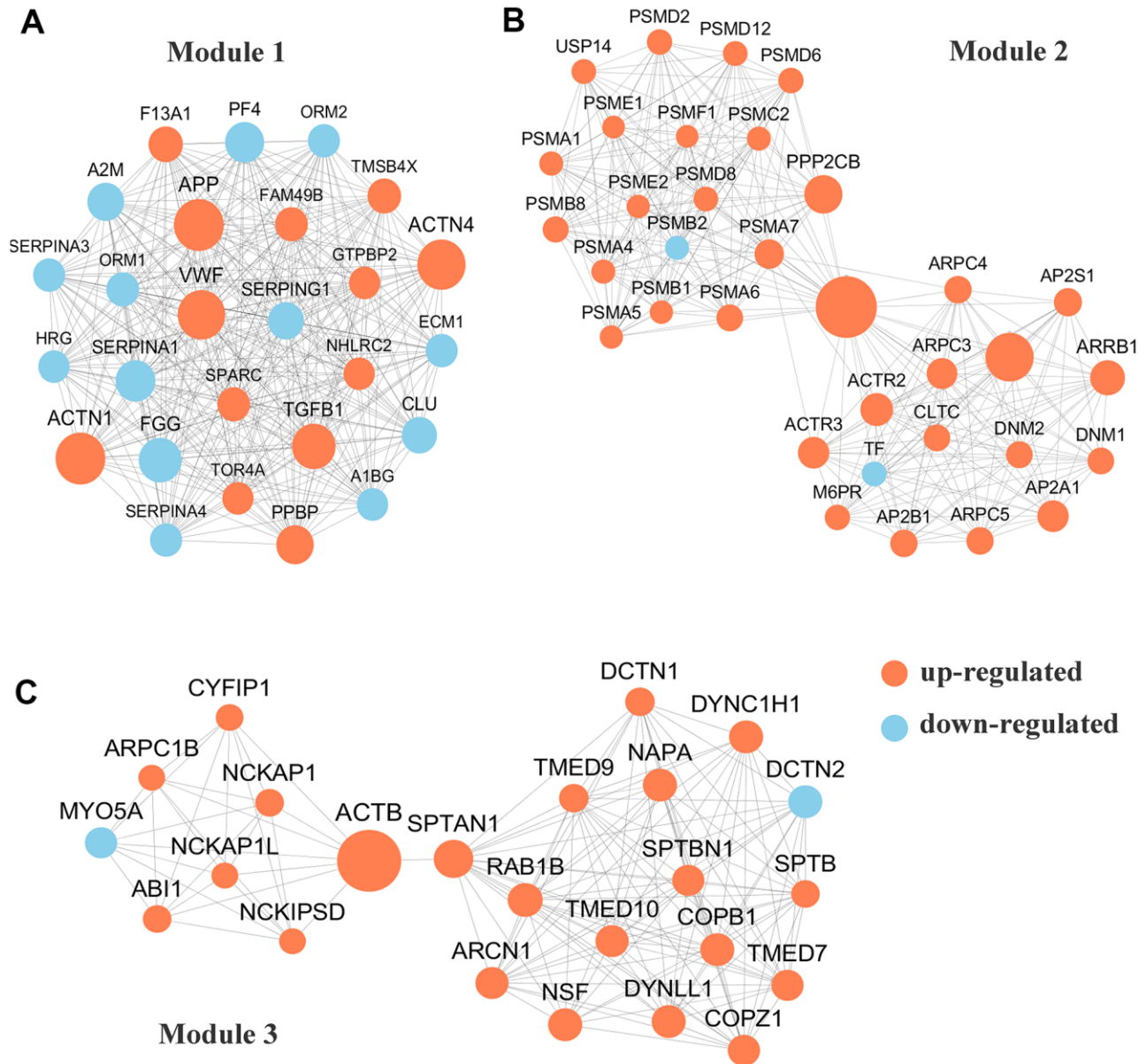


Figure 3. Protein-protein interaction (PPI) network analyses of differentially expressed proteins (DEPs) were performed and 3 most significant modules were yielded by molecular complex detection (MCODE) algorithm. Red and light cyan indicate up- and down-regulated differentially expressed proteins (DEPs), respectively. (A). Module 1 (score =33.000) was constructed with 26 nodes and 325 edges. (B) Module 2 (score =17.758) was constructed with 34 nodes and 293 edges. (C) Module 3 (score =12.261) was constructed with 24 nodes and 141 edges.

activated protein kinase binding for the up-regulated phosphoproteins. For the down-regulated phosphoproteins, calcium channel regulator activity, protein complex binding, macromolecular complex binding, and cell adhesion molecule binding were highly represented. Accordingly, in the analysis of cellular component enrichment classification, the highly enriched categories included the adherens junction, anchoring junction, cell-substrate adherens junction, focal adhesion, cell-substrate junction, secretory vesicle, and transport vesicle categories for the up-regulated phosphoproteins. For the down-regulated phosphoproteins, the cytoskeleton, vesicle, contractile fiber part, myofibril and cell junction categories were highly represented.

Site occupancies of phosphorylation correlate with the cellular signaling state [15]. A KEGG pathway analysis was performed on the DPPs to evaluate the pathways that were significantly represented in pSS. In the KEGG enrichment analysis of the phosphoproteins, there was no pathway which would represent up-regulated phosphoproteins. Compared with the proteome data, the phosphoproteome data showed the down-regulated phosphorylation of the signaling pathways in pSS, including the pathways of gap junctions, focal adhesion, phagosomes, the VEGF signaling pathway, and the inflammatory mediator regulation of TRP channels. Genes encoding interepithelial junctional proteins in adherens and gap junctions suggested perturbations in the permeability between epithelial barriers. Focal adhesion was associated with infiltrating leukocytes and was reported as an initial biomarker of possible autoimmunity, for the first time in 2011 in the lacrimal glands of mice [33].

Motif analysis of the phosphosites

Protein phosphorylation modifications are regulated by protein kinases (PKs), and different PKs prefer specific substrates with conserved motifs. We took advantage of the large number of phosphorylation sites identified in this study to carry out a bioinformatics analysis to identify novel phosphorylation motifs. We carried out an intensive sequence analysis for overrepresented phosphorylation site motifs surrounding the identified phosphosites (10 amino acids upstream and 10 amino acids downstream of each phosphosite) of the serine and threonine residues using the Motif-X program.

We identified 19 conserved motifs based on 1787 phospho-serine (pS) phosphosites, and 4 conserved motifs were identified based on 130 phospho-threonine (pT) phosphosites (Figure 4A). Because of the small number in the tyrosine group, no obvious conserved motif was acquired through our sequence analysis. In

particular, [xxxRxx_pS_PxPxxx] (motif 1), [RRx_pS_xxxxxx] (motif 6), [xxxxxx_pS_DxExxx] (motif 8), and [xxxxxx_pS_ExExxx] (motif 12) (motif score > 30.00) were strikingly conserved. Intriguingly, the motif analysis of the corresponding phosphosites found that the proline-directed motif [pS/pT_P] was increased significantly, whereas the most common motif, [xxxRxx_pS_PxPxxx] (motif 1), was more statistically significant and enriched, with a motif score of 41.28 and a fold increase of 33.1. According to the Human Protein Reference Database (HPRD), 6 phosphorylation motifs were experimentally verified and specific for known PKs, whereas 17 had no known corresponding kinases [34]. The detailed information and putative associated kinases are shown in Table 3. Motifs [xxxxxx_pS_DxExxx] (motif 8) and [xxxRxx_pT_xxxxxx] (motif 22) were verified as casein kinase II (CKII) substrate motifs [35]. CKII is a ubiquitous eukaryotic kinase [36] and phosphorylates numerous substrates, including many transcription factors [37, 38]. Phosphorylation by CKII has been associated with direct replication defect [39], which may be a possible explanation for the differentiation found in the PBMC phosphoproteome of pSS. Motif [xxxxxx_pT_Pxxxxx] (motif 21) was verified as a WW domain-binding motif [40], which mediates protein-protein interactions via proline-rich regions [41].

A heat map (Figure 4B) was generated to show the enrichment or depletion of specific amino acids neighboring the serine phosphosites. The amino acids aspartate (D), glutamate (E), glycine (G), proline (P) and arginine (R) had a tendency to be present in the proximity of serine phosphosites. The amino acids proline (P) and arginine (R) were greatly represented in the areas proximal to threonine phosphosites (Figure 4C). Arginine (R) was greatly represented in the sequence surrounding phosphosites but was greatly depleted in the -1 position. Cysteine (C), phenylalanine (F) and leucine (L) were notably depleted in the sequence surrounding phosphosites. These preferential amino acids near phosphosites reflect the specific recognition of the enzymes that catalyze phosphorylation in the PBMCs of pSS patients. Further studies are needed to investigate whether the different types of enzymes and kinases that regulate phosphorylation are active in the pathophysiology of pSS.

PPI and module analysis of the phosphoproteins

We selected the top 50 phosphoproteins with the closest interactions and generated a network with a total number of 58 associated phosphoproteins (Figure 5A). HSP90AA1, with 17 nodes, exhibited the most interactions among the phosphoproteins and was considered a hub phosphoprotein based on degree scores, which was valuable for other analyses in this

Table 3. Phosphorylation motifs enriched by motif-x and putative protein kinases.

Motif No.	Motif	Motif Score	Fold Increase	Putative Protein Kinases
Serine motif				
1	xxxRxx_S_PxPxxx	41.28	33.1	Novel
2	xxxxxx_S_PxxxRx	29.46	9.7	Novel
3	xxxRxx_S_Pxxxxx	25.83	9.1	Novel
4	xxxxxx_S_PxxxxK	23.79	8.7	Novel
5	xxxxxx_S_Pxxxxx	16.00	3.5	Novel
6	xxxRRx_S_xxxxxx	30.52	11.7	Novel
7	xxxRxx_S_xSxxxx	22.42	7.6	Novel
8	xxxxxx_S_DxExxx	32.00	13.7	Casein kinase II substrate motif
9	xxxRxx_S_xDxxxx	22.01	10.4	Novel
10	xxxxxx_S_xEDxxx	24.68	7.1	Novel
11	xxxRxx_S_xxxxxx	16.00	3.9	Calmodulin-dependent protein kinase II substrate motif
12	xxxxxx_S_ExExxx	32.00	9.5	Novel
13	xxxxxx_S_xxDxxx	14.58	2.4	Pyruvate dehydrogenase kinase substrate motif
14	xxxxxx_S_xExxxx	16.00	2.3	Novel
15	xxxxRx_S_xxxxxx	11.21	2.3	Novel
16	xxxxxx_S_xDxxxx	9.18	2.2	Novel
17	xxxxxG_S_xxxxxx	7.48	2.0	Novel
18	xxxxxx_S_Fxxxxx	6.77	2.3	Novel
19	xxxxxD_S_xxxxxx	6.53	2.2	β -Adrenergic Receptor kinase substrate motif
Threonine motif				
20	xxxRxx_T_Pxxxxx	26.69	24.6	Novel
21	xxxxxx_T_Pxxxxx	16.00	3.8	WW domain binding motif
22	xxxRxx_T_xxxxxx	6.83	3.1	Calmodulin-dependent protein kinase II substrate motif

study. Further analysis of the complex by MCODE revealed 2 subnetworks for the network Cluster 1 (MCODE score = 5.000) was constructed with 5 nodes and 10 edges (Figure 5B) and cluster 2 (MCODE score =3.000) was constructed with 3 nodes and 3 edges (Figure 5C).

CONCLUSION

Based on proteome and phosphoproteome analyses, our results provide new insights into pSS and reveal

many essential features of and differences in the PBMCS of pSS through a systematic bioinformatics analysis. In total, 787 proteins were identified as DEPs, and 175 phosphosites on 123 proteins were identified as DPPs. Using systematic bioinformatics analyses, we identified 10 hub proteins: *UBA52*, *MAPK1*, *HSPA8*, *ACTB*, *ACLY*, *APP*, *ACTN1*, *ACTN4*, *VWF*, and *TGFBI*. Finally, we identified 22 motifs using motif analysis of the phosphosites and found 17 newly identified motifs, while 6 motifs were experimentally verified for

known protein kinases. This allowed us to conclude that a panel of candidate biomarkers rather than a single specific protein is apparently able to best distinguish pSS among individuals, including healthy volunteers, and when they have been further validated, these biomarkers may improve the clinical detection of primary SS.

The analyses of the DEPs and DPPs for which the expression may correlate with the molecular mechanisms of pSS provide novel insight into the roles of phosphorylation and the underlying cause or regulation of this disease. Additionally, the hub proteins, pathways and motifs identified might be

therapeutic targets and potential discriminatory biomarkers for pSS diagnosis.

While our results are promising, we are aware of the limitations of our study. Our study was restricted to cell-lined signatures and was not further validated in the clinical setting. The lack of multiple biological replicates requires further experiments to consolidate the conclusions drawn from the proteomics study data. Future studies will test the ability of these biomarker levels, alone and in combination, to diagnose the PBMC component of pSS. Overall, this work provides a unique prospect and resource for future studies focusing on the PBMCs of pSS patients.

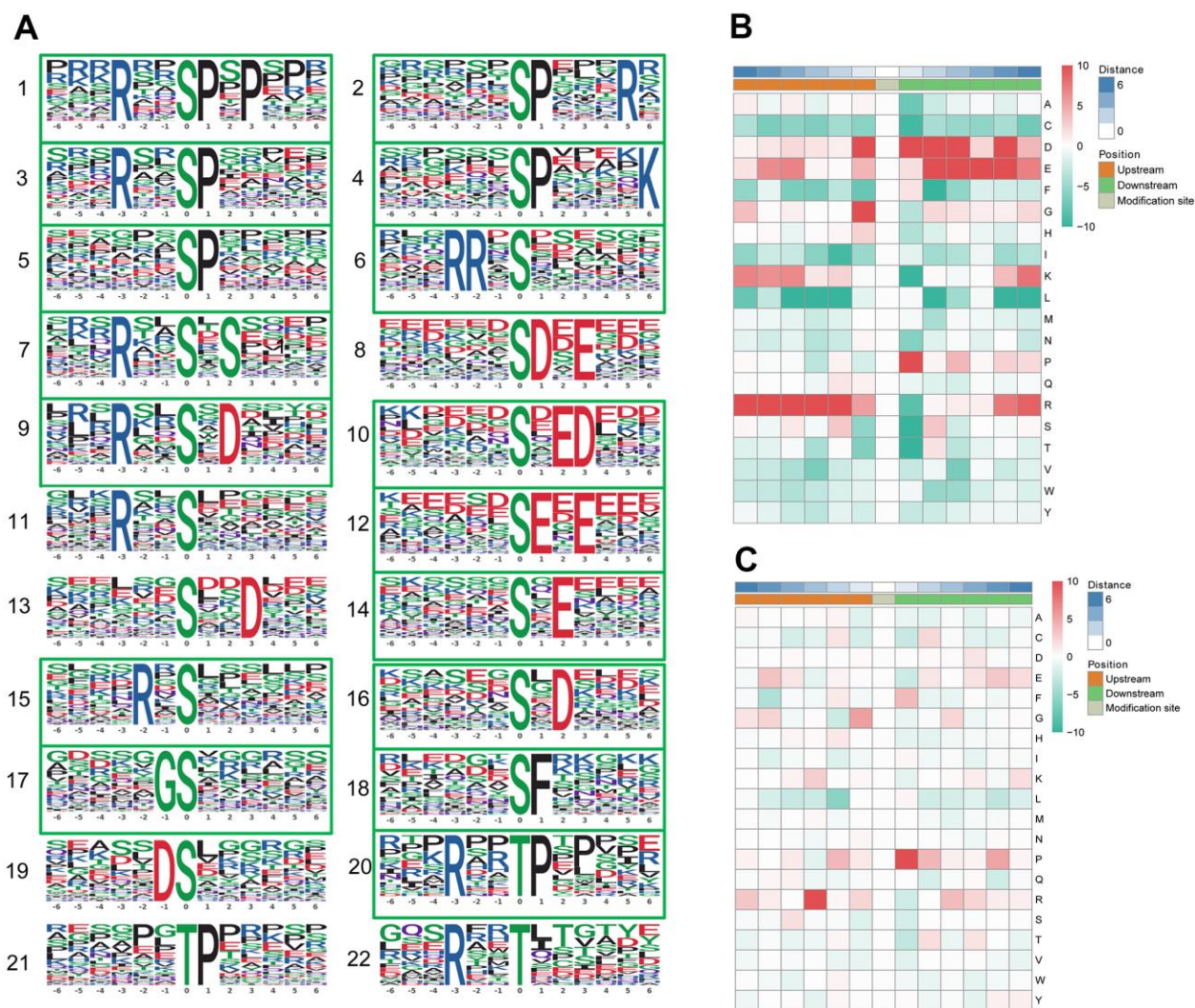


Figure 4. Motif analysis of the phosphosites. (A) Significantly enriched phosphorylation motifs extracted from the overrepresented phosphopeptide dataset by Motif-X. (1) -(19) Motifs from phosphoserine; (20) -(22) Motifs from phosphothreonine. Among these motifs, 6 are identified as 5 known phosphorylation motifs and 17 are newly identified (enclosed into green boxes). The detailed information and putative associated kinases are shown in Table 3. Motif enrichment heat map of phosphoserine (B) and phosphothreonine (C) upstream and downstream of all identified phosphorylation modification sites. Red indicates significant enrichment of the amino acid near the modification site, while green indicates significant reduction of the amino acid near the modification site.

MATERIALS AND METHODS

Patient assessment and PBMC isolation

In Supplementary Table 2, the general clinical characteristics of the study subjects are summarized. Eight pSS patients fulfilling the revised 2002 American-European Consensus Group criteria were enrolled in the study [9]. All patients had the clinical symptoms of dry eyes and/or mouth and at least one of the qualifying anti-SS-related antigen A (Ro/SSA), anti-SS-related antigen La/SSB, or anti-nuclear antibodies and rheumatoid factor. Patients with HIV-1, hepatitis B and/or hepatitis C infections were excluded from the study. None of the patients had received specific treatment, such as glucocorticoid and/or biological therapy and/or immunosuppressive drug treatment at the time their blood was drawn for this study. In addition, ten normal controls (NC) (between 20 and 65 years old)

who were enrolled in the Physical Examination Centre at the Shenzhen People's Hospital for routine health examinations and without disease were included. The healthy subjects had no complaints related to oral dryness and autoimmune disease and had normal salivary function. Written informed consent was obtained from all patients and healthy volunteers before they were included in the study.

Blood samples from all patients were obtained before breakfast. We then took 4mL of each sample and it was diluted with an equal volume of normal saline. Each of these samples was mixed and added to a 6mL lymphocyte separation solution. The samples were immediately processed by density gradient centrifugation according to the standardized protocol of PBMC isolation to minimize the degradation of proteins. The samples were stored at -80°C until they were used in the proteomic analyses.

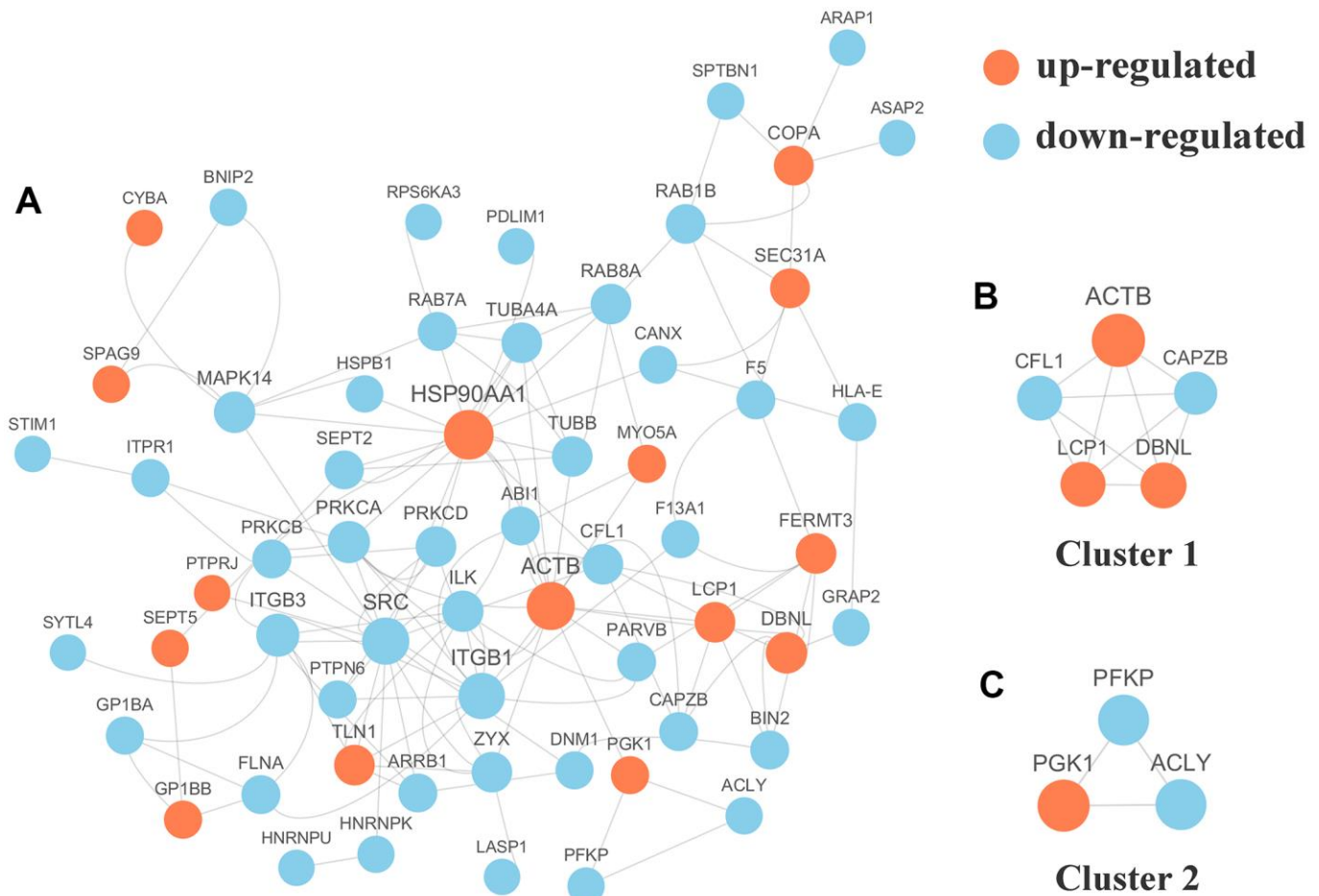


Figure 5. (A) Protein-protein interaction (PPI) network analyses of differentially expressed phosphorylated proteins (DPPs) were performed and 2 significant clusters were yielded by molecular complex detection (MCODE) algorithm. Red and light cyan indicate up- and down-regulated DPPs, respectively. Yellow indicates hub phosphorylated proteins based on degree. (B) Cluster 1 (MCODE score = 5.000) was constructed with 5 nodes and 10 edges. (C) Cluster 2 (MCODE score = 3.000) was constructed with 3 nodes and 3 edges.

Experimental procedures

Materials and reagents

The materials and reagents required for the sample preparation are shown in Supplementary Table 3.

Protein extraction

The PBMCs were removed from -80°C storage, and four volumes of lysis buffer (containing 8 mol/L urea, 1% protease inhibitors, and 1% phosphorylase inhibitor) was added to a high-intensity ultrasonic processor instrument (PTM Bio, Hangzhou, China), after which the residual cell components were removed immediately by centrifugation ($12000 \times g$ at 4°C for 10 min). Then, the supernatant was transferred to a new tube, and the protein concentration was defined by a BCA kit according to the manufacturer's instructions.

Trypsin digestion

The protein solution was sequentially diluted (5 mmol/L dithiothreitol for 30 min at 56°C) and alkylated with 11 mmol/L iodoacetamide for 15 min. These procedures were performed in darkness at room temperature. Then, the assembled protein sample was diluted to a urea concentration of less than 2 mol/L. Finally, trypsin was added to initiate overnight digestion (the ratio of trypsin to the protein mass ratio was 1:50) at 37°C and a subsequent 4 h digestion (the ratio of trypsin to protein mass was 1:100).

Affinity enrichment of the phosphopeptides

Peptide mixtures were first incubated with IMAC microsphere suspensions with vibration in loading buffer (50% acetonitrile/6% trifluoroacetic acid). The IMAC microspheres with enriched phosphopeptides were collected by centrifugation, and the supernatant was removed. To remove nonspecifically adsorbed peptides, the IMAC microspheres were washed sequentially with 50% acetonitrile/6% trifluoroacetic acid and 30% acetonitrile/0.1% trifluoroacetic acid sequentially. To elute the enriched phosphopeptides from the IMAC microspheres, an elution buffer containing 10% NH_4OH was added, and the enriched phosphopeptides were eluted. Finally, the eluted fractions were combined and vacuum-dried. For LC-MS/MS analyses, the resulting peptides were desalted with C18 ZipTips (Millipore) according to the manufacturer's instructions.

LC-MS/MS analysis

The tryptic peptides were dissolved in an aqueous solution containing 0.1% formic acid (solvent A), and a NanoElite ultrahigh-performance liquid system was used for separation after dissolution. The gradient was composed of solvent B (0.1% formic acid in 98% acetonitrile) increased from 2% to 22% over 70 min for

the proteome or 40 min for the phosphoproteome and from 22% to 37% (22% to 35% for the phosphoproteome) over 10 min and increasing to 80% over 5 min, then held at 80% for the final 5 min. For all steps, the flow rate was maintained at a constant 300 nL/min for the proteome and 250 nL/min for the phosphoproteome in an EASY-nLC 1000 UPLC system. The peptide was separated by an ultrahigh-performance liquid system, subjected to a capillary ion source for ionization, and then analyzed by timsTOF Pro mass spectrometry. The electrospray voltage was applied at 1.4 kV. TOF mass spectrometry was used to detect and analyze the peptide ions and their secondary fragments. The m/z scan range was 100 to 1700 for the secondary mass spectrometer. The data acquisition mode was based on the parallel cumulative serial fragmentation (PASEF) mode. After the acquisition of the first-level mass spectrometry data, the 10-fold PASEF mode was used to collect the second-level spectrum of the parent ion with a charge number in the range of 0-5. The dynamic exclusion time of tandem mass spectrometer scanning is set to 30 seconds to avoid repeated scanning of the parent ion.

Database search

The resulting MS/MS data were processed using the MaxQuant search engine (v1.6.6.0). The retrieval parameter settings were as follows: the database was homo_sapiens_9606_SP_20190513 with 20422 sequences; an anti-library was added to calculate the FDR caused by random matches, and a common pollution library was added to the database to eliminate the influence of contaminated proteins in the identification results; the digestion method was set to Trypsin/P; the number of missing cut points was set to 2; the minimum length of the peptide was set to 7 amino acid residues; the maximum number of modifications of the peptide was set to 5; the tolerance values of the mass error for the primary precursor ion in the first search and main search were set to 40 ppm and 40 ppm, respectively, and the error tolerance of the mass of the secondary fragment ion was to 0.04 Da; the cysteine alkylation was specified as a fixed modification, and the oxidation of methionine and the acetylation of the N-terminus of the protein were specified as the variable modifications; and the FDR for the protein identification and peptide-spectrum matches identification was adjusted to 1%. The mass tolerance for the precursor ions was set as 20 ppm for the first search and 5 ppm in the main search, and the mass tolerance for the fragment ions was set as 0.02 Da. Carbamidomethyl on Cys was specified as a fixed modification, and acetylation modification and oxidation on Met were specified as variable modifications. The FDR was adjusted to $< 1\%$, and the minimum score for modified peptides was set to be > 40 . Only proteins with a fold change > 2 or < 0.5 and a

p value < 0.01 were considered to exhibit a significant difference and were subjected to the subsequent bioinformatics analyses.

Quality control

Most of the peptides were distributed between 7-20 amino acids in length, which conformed to the general rules based on trypsin enzymatic hydrolysis and met the requirements of quality control (Supplementary Figure 6A). Proteins below 10KD (which are more peptide-oriented) were lost due to multiple acetone precipitation and rinsing; the distribution of proteins above 10KD was relatively uniform, indicating that there was no significant molecular weight bias for proteins above 10KD in the sample preparation process. The protein above 100KD was not lost during the preparation process due to poor solubility (Supplementary Figure 6B). The coverage of most proteins was below 20%. In the mass spectrometry method based on the shotgun (also called bottom-up) strategy, the mass spectrometer preferentially scanned the peptides with higher abundance. Therefore, the coverage of the protein was positively correlated with its abundance in the sample (Supplementary Figure 6C). The molecular weight of the protein had a negative correlation with the coverage. The reason was that a protein with a larger molecular weight could theoretically produce more enzymatically digested peptides. To achieve the same coverage, large proteins required more peptides to be identified (Supplementary Figure 7A). The majority of the phosphorylated proteins carried 1 to 4 Khib sites, and approximately 6% of the proteins carried more than 4 Khib sites (Supplementary Figure 7B).

Bioinformatics methods

For the annotation analyses, GO annotation of the proteome was based on the UniProt-GOA database (<http://www.ebi.ac.uk/GOA/>) with the IDs of the proteins found in this study of pSS; these proteins were then classified by GO annotation based on the following three categories: biological process, molecular function and cellular component. The functional protein domain descriptions were annotated based on the InterPro domain database (<http://www.ebi.ac.uk/interpro/>) and the protein sequence alignment. We also conducted KEGG pathway annotation using the KEGG database (<http://www.genome.jp/kegg/>), and WoLF PSORT (v.0.2 http://www.genscript.com/psort/wolf_psort.html) was used to predict and characterize the subcellular localization of the identified proteins. For the functional enrichment analyses of the proteins identified by GO, a two-tailed Fisher's exact test was employed as a significance test to verify the enrichment of the DEPs for each GO annotation category in our

study (p value < 0.01). For the functional enrichment analyses of the proteins identified in KEGG pathways, the KEGG database was used to identify the pathways enriched with the proteins and to test the enrichment of the DEPs compared to all the identified proteins. A two-tailed Fisher's exact test was used, and the pathways with corrected p value < 0.01 were considered significant and were separated into individual categories. For the enrichment analyses of the protein domains, a two-tailed Fisher's exact test was utilized to examine the enrichment of the DEPs compared to all the identified proteins. For enrichment-based clustering, we first collated all the categories based on the results from both enrichment analyses along with the respective p values for clustering analyses depending on the DEP functional classification (based on the GO, domain and pathway analyses). Then, we picked the categories with at least one enriched cluster and a p value < 0.01. For the PPI network, the STRING database (version 10.5) was used, and a confidence score was defined to assess the interactions, with a confidence score >0.7 defined as a high confidence score. Finally, we integrated the databases and networks and used the MCODE algorithm to analyze the characteristics of the networks and find densely connected regions. For phospho-proteome, software MoMo (motif-x algorithm) was used to analyze the model of sequences constituted with amino acids in specific positions of modify-21-mers (10 amino acids upstream and downstream of the site, but phosphorylation with modify-13-mers that 6 amino acids upstream and downstream of the site) in all protein sequences. Only when the minimum number of occurrences was set to 20 and the statistical test P value is less than 0.000001, the characteristic sequence form is considered to be a motif of the modified peptide. Emulate original motif-x was ticked, and other parameters with default.

Abbreviations

pSS: primary Sjögren syndrome; AECG: American-European classification criteria; PBMCs: peripheral blood mononuclear cells; NK: natural killer; DEPs: differentially expressed proteins; DPPs: differentially phosphorylated proteins; IMAC: immobilized metal affinity chromatography; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interaction; NC: normal control; PASEF: parallel cumulative serial fragmentation; FDR: false positive rate; MCODE: molecular complex detection; ARVC: arrhythmogenic right ventricular cardiomyopathy; HCM: hypertrophic cardiomyopathy; PKs: protein kinases; pS: phospho-serine; pT: phospho-threonine; HPRD: Human Protein Reference Database; CKII: casein kinase II.

AUTHOR CONTRIBUTIONS

Shaoying Huang: writing- original draft preparation, visualization, software. Fengping Zheng: writing- reviewing and editing. Lixiong Liu: data curation, investigation. Shuhui Meng: investigation, software. Wanxia Cai: investigation, supervision. Cantong Zhang: investigation. Weier Dai: methodology, validation. Dongzhou Liu: methodology. Xiaoping Hong: conceptualization, methodology. Donge Tang: conceptualization, methodology, validation. Yong Dai: conceptualization, methodology.

ACKNOWLEDGMENTS

We thank Andong He for his assistance with the algorithms in PPI network analyses. We thank the study participants without whom this study would not be possible.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China [Grant No. 81671596]; the Natural Science Foundation of Guangxi [Grant No. 2017GXNSFA A198375]; the National Science Foundation for Young Scientists of China [Grant No. 31700795]; the Key Research and Development Program of Guangdong Province [Grant No. 2019B020229001]; and the National Natural Science Foundation of China [Grant No. 81971464]; and Shenzhen Key Medical Discipline Construction Fund [Grant No.SZXXK011]; Guangxi Key Laboratory of Metabolic Diseases Research [Grant No.20-065-76].

REFERENCES

1. Fox RI. Sjögren's syndrome. *Lancet*. 2005; 366:321–31. [https://doi.org/10.1016/S0140-6736\(05\)66990-5](https://doi.org/10.1016/S0140-6736(05)66990-5) PMID:16039337
2. Ramos-Casals M, Brito-Zerón P, Seror R, Bootsma H, Bowman SJ, Dörner T, Gottenberg JE, Mariette X, Theander E, Bombardieri S, De Vita S, Mandl T, Ng WF, et al, and EULAR Sjögren Syndrome Task Force. Characterization of systemic disease in primary Sjögren's syndrome: EULAR-SS Task Force recommendations for articular, cutaneous, pulmonary and renal involvements. *Rheumatology (Oxford)*. 2015; 54:2230–8. <https://doi.org/10.1093/rheumatology/kev200> PMID:26231345
3. Ramos-Casals M, Brito-Zerón P, Sisó-Almirall A, Bosch X. Primary Sjogren syndrome. *BMJ*. 2012; 344:e3821.

- <https://doi.org/10.1136/bmj.e3821> PMID:22700787
4. Lin G, Wang J, Ge J, eds. *Practice of Internal Medicine*. 15th ed. Beijing: People's Medical Publishing House; 2017. 2647p (<https://www.amazon.cn/dp/B07V7YDMWS>)
5. Alamanos Y, Tsifetaki N, Voulgari PV, Venetsanopoulou AI, Siozos C, Drosos AA. Epidemiology of primary Sjögren's syndrome in north-west Greece, 1982-2003. *Rheumatology (Oxford)*. 2006; 45:187–91. <https://doi.org/10.1093/rheumatology/kei107> PMID:16332955
6. Gøransson LG, Haldorsen K, Brun JG, Harboe E, Jonsson MV, Skarstein K, Time K, Omdal R. The point prevalence of clinically relevant primary Sjögren's syndrome in two Norwegian counties. *Scand J Rheumatol*. 2011; 40:221–24. <https://doi.org/10.3109/03009742.2010.536164> PMID:21231797
7. Zhang NZ, Shi CS, Yao QP, Pan GX, Wang LL, Wen ZX, Li XC, Dong Y. Prevalence of primary Sjogren's syndrome in China. *J Rheumatol*. 1995; 22:659–61. PMID:7791159
8. Saraux A, Pers JO, Devauchelle-Pensec V. Treatment of primary Sjögren syndrome. *Nat Rev Rheumatol*. 2016; 12:456–71. <https://doi.org/10.1038/nrrheum.2016.100> PMID:27411907
9. Vitali C, Bootsma H, Bowman SJ, Dorner T, Gottenberg JE, Mariette X, Ramos-Casals M, Ravaud P, Seror R, Theander E, Tzioufas AG. Classification criteria for Sjogren's syndrome: we actually need to definitively resolve the long debate on the issue. *Ann Rheum Dis*. 2013; 72:476–78. <https://doi.org/10.1136/annrheumdis-2012-202565> PMID:23264355
10. Harley JB, Scofield RH. Systemic lupus erythematosus: RNA-protein autoantigens, models of disease heterogeneity, and theories of etiology. *J Clin Immunol*. 1991; 11:297–316. <https://doi.org/10.1007/BF00918796> PMID:1722216
11. Hendrick JP, Wolin SL, Rinke J, Lerner MR, Steitz JA. Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol Cell Biol*. 1981; 1:1138–49. <https://doi.org/10.1128/mcb.1.12.1138> PMID:6180298
12. Hu S, Wang J, Meijer J, Jeong S, Xie Y, Yu T, Zhou H, Henry S, Vissink A, Pijpe J, Kallenberg C, Elashoff D, Loo JA, Wong DT. Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. *Arthritis Rheum*. 2007; 56:3588–600.

<https://doi.org/10.1002/art.22954>

PMID:17968930

13. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, Daniels TE, Fox PC, Fox RI, Kassin SS, Pillemer SR, Talal N, Weisman MH, and European Study Group on Classification Criteria for Sjögren's Syndrome. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis*. 2002; 61:554–58.
<https://doi.org/10.1136/ard.61.6.554>
PMID:12006334
14. Cohen P. Signal integration at the level of protein kinases, protein phosphatases and their substrates. *Trends Biochem Sci*. 1992; 17:408–13.
[https://doi.org/10.1016/0968-0004\(92\)90010-7](https://doi.org/10.1016/0968-0004(92)90010-7)
PMID:1333658
15. Sharma K, D'Souza RC, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep*. 2014; 8:1583–94.
<https://doi.org/10.1016/j.celrep.2014.07.036>
PMID:25159151
16. Fogdell-Hahn A, Ligiers A, Grønning M, Hillert J, Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens*. 2000; 55:140–48.
<https://doi.org/10.1034/j.1399-0039.2000.550205.x>
PMID:10746785
17. Nakamura J, Meguro A, Ishii G, Mihara T, Takeuchi M, Mizuki Y, Yuda K, Yamane T, Kawagoe T, Ota M, Mizuki N. The association analysis between HLA-A*26 and Behçet's disease. *Sci Rep*. 2019; 9:4426.
<https://doi.org/10.1038/s41598-019-40824-y>
PMID:30872678
18. Tholander F, Muroya A, Roques BP, Fournié-Zaluski MC, Thunnissen MM, Haeggström JZ. Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chem Biol*. 2008; 15:920–29.
<https://doi.org/10.1016/j.chembiol.2008.07.018>
PMID:18804029
19. Moroz OV, Burkitt W, Wittkowski H, He W, Ianoul A, Novitskaya V, Xie J, Polyakova O, Lednev IK, Shekhtman A, Derrick PJ, Bjoerk P, Foell D, Bronstein IB. Both Ca²⁺ and Zn²⁺ are essential for S100A12 protein oligomerization and function. *BMC Biochem*. 2009; 10:11.
<https://doi.org/10.1186/1471-2091-10-11>
PMID:19386136
20. Stewart CR, Stuart LM, Wilkinson K, van Gils JM, Deng J, Halle A, Rayner KJ, Boyer L, Zhong R, Frazier WA, Lacy-Hulbert A, El Khoury J, Golenbock DT, Moore KJ. CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat Immunol*. 2010; 11:155–61.
<https://doi.org/10.1038/ni.1836> PMID:20037584
21. Tsui C, Martinez-Martin N, Gaya M, Maldonado P, Llorian M, Legrave NM, Rossi M, MacRae JI, Cameron AJ, Parker PJ, Leitges M, Bruckbauer A, Batista FD. Protein kinase c-β dictates B cell fate by regulating mitochondrial remodeling, metabolic reprogramming, and heme biosynthesis. *Immunity*. 2018; 48:1144–59.e5.
<https://doi.org/10.1016/j.immuni.2018.04.031>
PMID:29884460
22. Alunno A, Ibba-Manneschi L, Bistoni O, Cipriani S, Topini F, Gerli R, Manetti M. Angiogenic T cells in primary Sjogren's syndrome: a double-edged sword? *Clin Exp Rheumatol*. 2019 (Suppl 118); 37:36–41.
PMID:30336792
23. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol*. 2010; 6:e1000641.
<https://doi.org/10.1371/journal.pcbi.1000641>
PMID:20090828
24. Zaki N, Berenguères J, Efimov D. Detection of protein complexes using a protein ranking algorithm. *Proteins*. 2012; 80:2459–68.
<https://doi.org/10.1002/prot.24130> PMID:22685080
25. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015; 43:D447–52.
<https://doi.org/10.1093/nar/gku1003>
PMID:25352553
26. Hou H, Li D, Yao YH, Lu J, Sun YN, Hu YX, Wu SY, Chu XR, Xiao PF, Xu GQ, Hu SY. Proteomic analysis for identifying the differences in molecular profiling between fanconi anaemia and aplastic anaemia. *Am J Transl Res*. 2019; 11:6522–33.
PMID:31737203
27. Davis RJ. The mitogen-activated protein kinase signal transduction pathway. *J Biol Chem*. 1993; 268:14553–56.
PMID:8325833
28. Avruch J, Zhang XF, Kyriakis JM. Raf meets Ras: completing the framework of a signal transduction pathway. *Trends Biochem Sci*. 1994; 19:279–83.
[https://doi.org/10.1016/0968-0004\(94\)90005-1](https://doi.org/10.1016/0968-0004(94)90005-1)
PMID:8048167

29. Crawford TQ, Ndhlovu LC, Tan A, Carvidi A, Hecht FM, Sinclair E, Barbour JD. HIV-1 infection abrogates CD8+ T cell mitogen-activated protein kinase signaling responses. *J Virol*. 2011; 85:12343–50. <https://doi.org/10.1128/JVI.05682-11> PMID:[21937661](https://pubmed.ncbi.nlm.nih.gov/21937661/)
30. Zhao B, Xu S, Dong X, Lu C, Springer TA. Prodomain-growth factor swapping in the structure of pro-TGF- β 1. *J Biol Chem*. 2018; 293:1579–89. <https://doi.org/10.1074/jbc.M117.809657> PMID:[29109152](https://pubmed.ncbi.nlm.nih.gov/29109152/)
31. Kotlarz D, Marquardt B, Barøy T, Lee WS, Konnikova L, Hollizeck S, Magg T, Lehle AS, Walz C, Borggraefe I, Hauck F, Bufler P, Conca R, et al. Human TGF- β 1 deficiency causes severe inflammatory bowel disease and encephalopathy. *Nat Genet*. 2018; 50:344–48. <https://doi.org/10.1038/s41588-018-0063-6> PMID:[29483653](https://pubmed.ncbi.nlm.nih.gov/29483653/)
32. Hebert AS, Dittenhafer-Reed KE, Yu W, Bailey DJ, Selen ES, Boersma MD, Carson JJ, Tonelli M, Balloon AJ, Higbee AJ, Westphall MS, Pagliarini DJ, Prolla TA, et al. Calorie restriction and SIRT3 trigger global reprogramming of the mitochondrial protein acetylome. *Mol Cell*. 2013; 49:186–99. <https://doi.org/10.1016/j.molcel.2012.10.024> PMID:[23201123](https://pubmed.ncbi.nlm.nih.gov/23201123/)
33. Peck AB, Saylor BT, Nguyen L, Sharma A, She JX, Nguyen CQ, McIndoe RA. Gene expression profiling of early-phase Sjögren's syndrome in C57BL/6.NOD-Aec1Aec2 mice identifies focal adhesion maturation associated with infiltrating leukocytes. *Invest Ophthalmol Vis Sci*. 2011; 52:5647–55. <https://doi.org/10.1167/iovs.11-7652> PMID:[21666236](https://pubmed.ncbi.nlm.nih.gov/21666236/)
34. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, et al. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009; 37:D767–72. <https://doi.org/10.1093/nar/gkn892> PMID:[18988627](https://pubmed.ncbi.nlm.nih.gov/18988627/)
35. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol*. 2005; 23:1391–98. <https://doi.org/10.1038/nbt1146> PMID:[16273072](https://pubmed.ncbi.nlm.nih.gov/16273072/)
36. Heyen BJ, Alsheikh MK, Smith EA, Torvik CF, Seals DF, Randall SK. The calcium-binding activity of a vacuole-associated, dehydrin-like protein is regulated by phosphorylation. *Plant Physiol*. 2002; 130:675–87. <https://doi.org/10.1104/pp.002550> PMID:[12376635](https://pubmed.ncbi.nlm.nih.gov/12376635/)
37. Allende JE, Allende CC. Protein kinases. 4. Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. *FASEB J*. 1995; 9:313–23. <https://doi.org/10.1096/fasebj.9.5.7896000> PMID:[7896000](https://pubmed.ncbi.nlm.nih.gov/7896000/)
38. Issinger OG. Casein kinases: pleiotropic mediators of cellular regulation. *Pharmacol Ther*. 1993; 59:1–30. [https://doi.org/10.1016/0163-7258\(93\)90039-g](https://doi.org/10.1016/0163-7258(93)90039-g) PMID:[8259381](https://pubmed.ncbi.nlm.nih.gov/8259381/)
39. McShan GD, Wilson VG. Contribution of bovine papillomavirus type 1 E1 protein residue 48 to replication function. *J Gen Virol*. 2000; 81:1995–2004. <https://doi.org/10.1099/0022-1317-81-8-1995> PMID:[10900038](https://pubmed.ncbi.nlm.nih.gov/10900038/)
40. Tzivion G, Shen YH, Zhu J. 14-3-3 proteins; bringing new definitions to scaffolding. *Oncogene*. 2001; 20:6331–38. <https://doi.org/10.1038/sj.onc.1204777> PMID:[11607836](https://pubmed.ncbi.nlm.nih.gov/11607836/)
41. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P. Characterization of a novel protein-binding module—the WW domain. *FEBS Lett*. 1995; 369:67–71. [https://doi.org/10.1016/0014-5793\(95\)00550-s](https://doi.org/10.1016/0014-5793(95)00550-s) PMID:[7641887](https://pubmed.ncbi.nlm.nih.gov/7641887/)

SUPPLEMENTARY MATERIALS

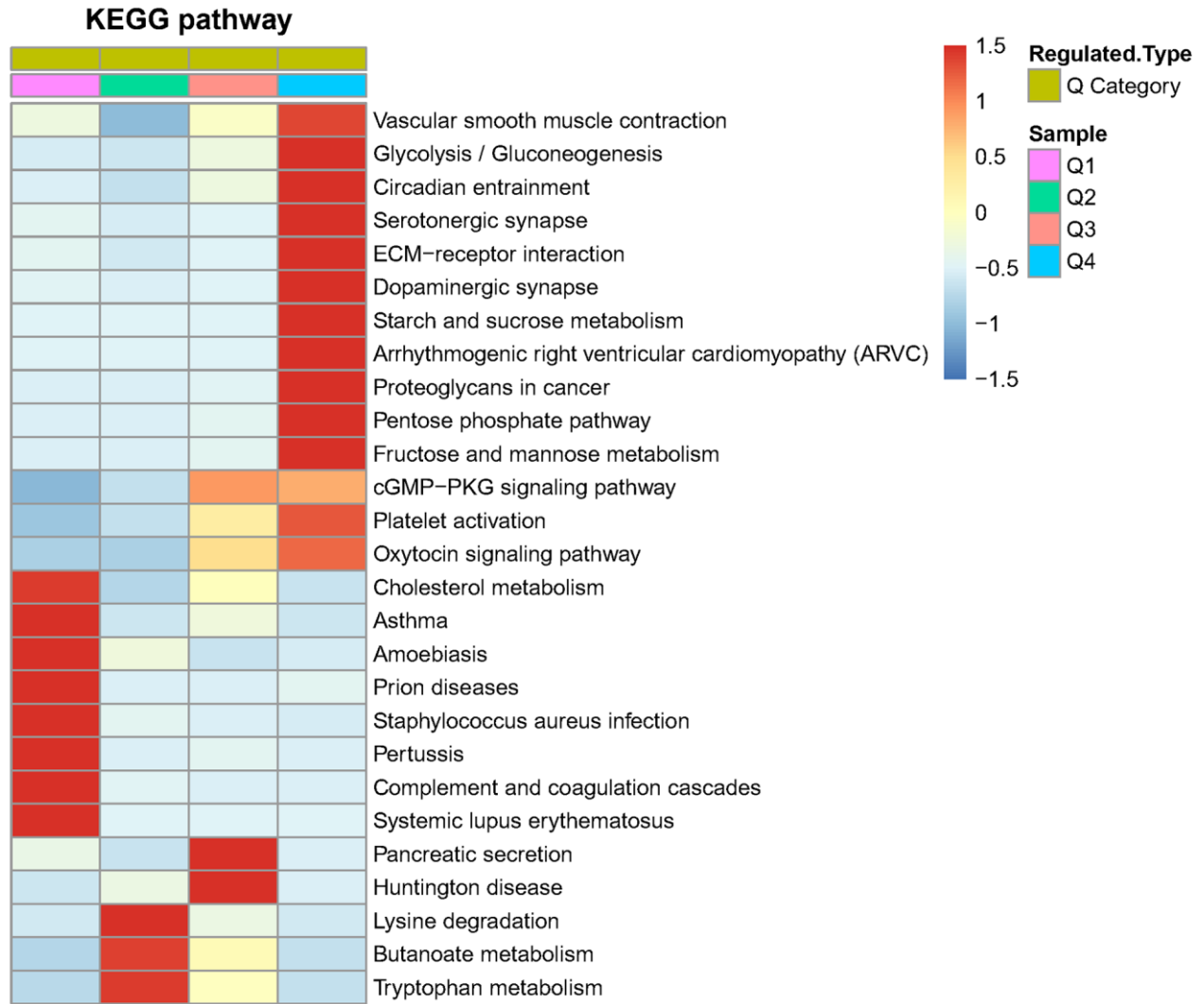
Supplementary Figures



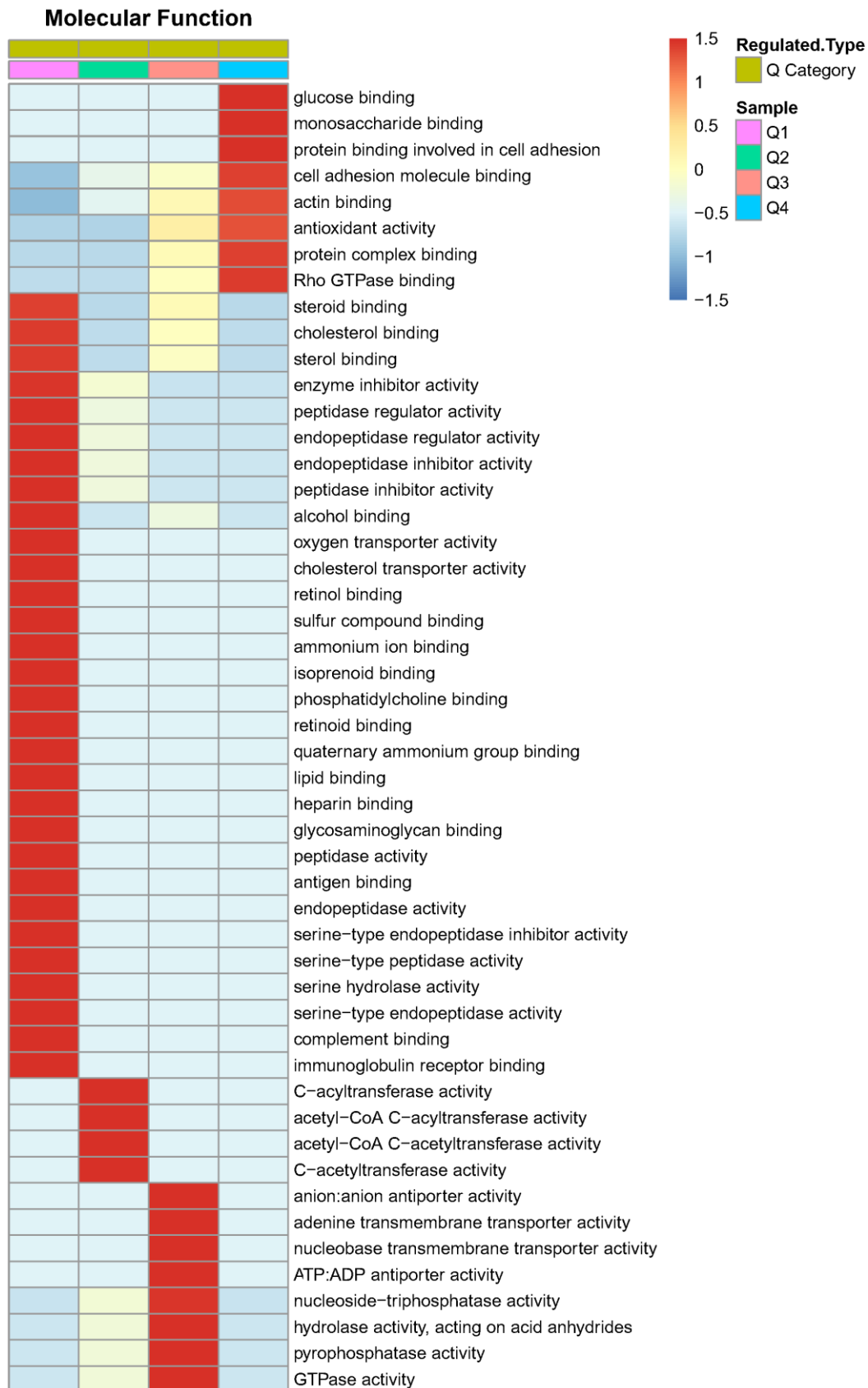
Supplementary Figure 1. Functional enrichment clustering analyses of differentially expressed proteins (DEPs) in four quantiles (Q1-Q4) based on biological process.



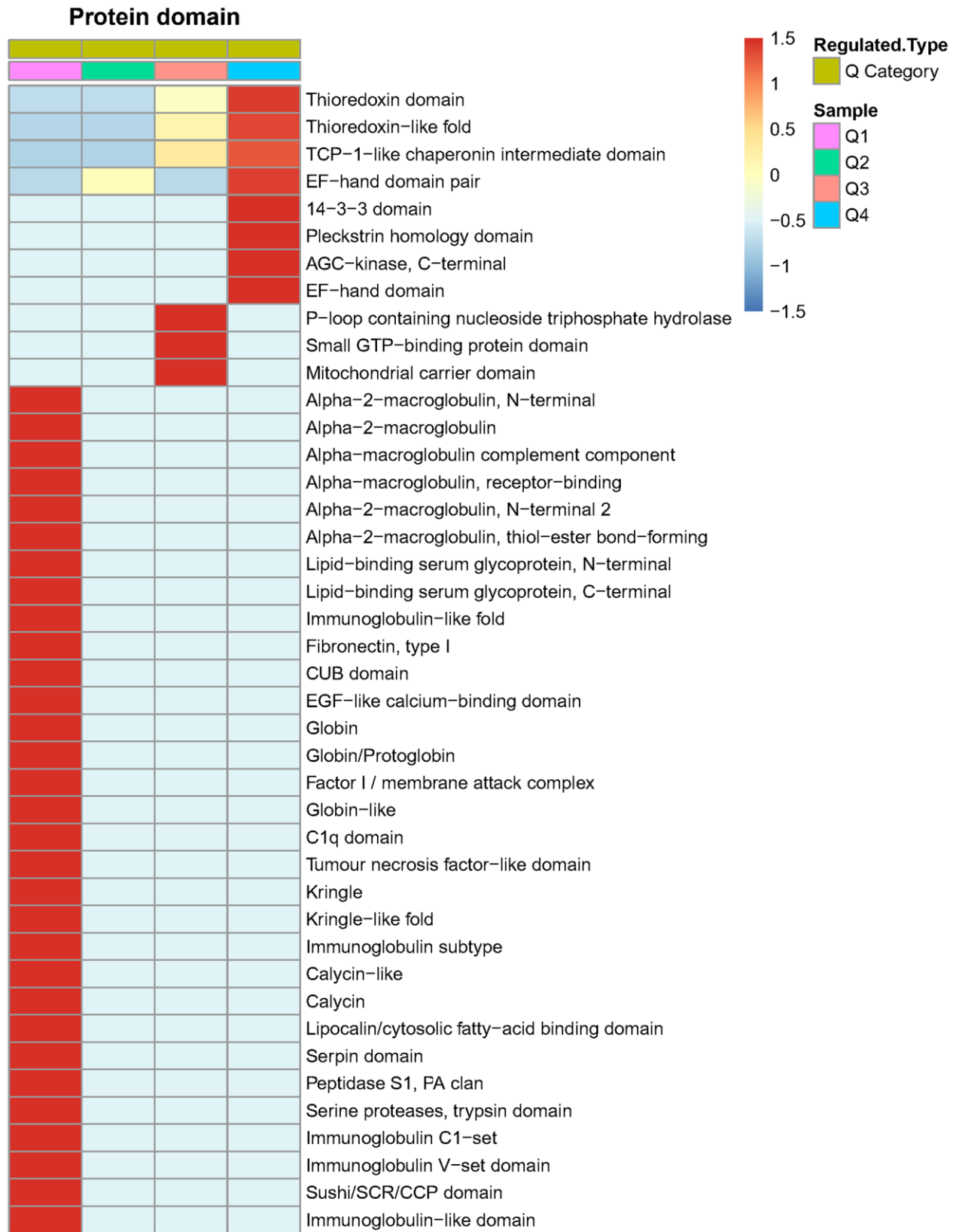
Supplementary Figure 2. Functional enrichment clustering analyses of differentially expressed proteins (DEPs) in four quantiles (Q1-Q4) based on cellular component.



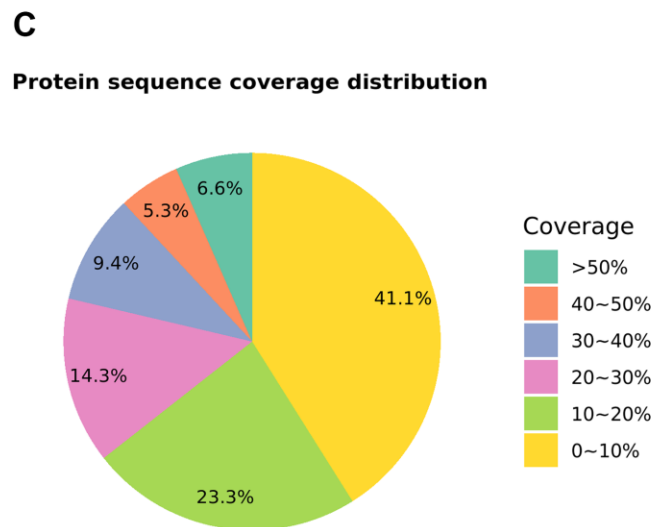
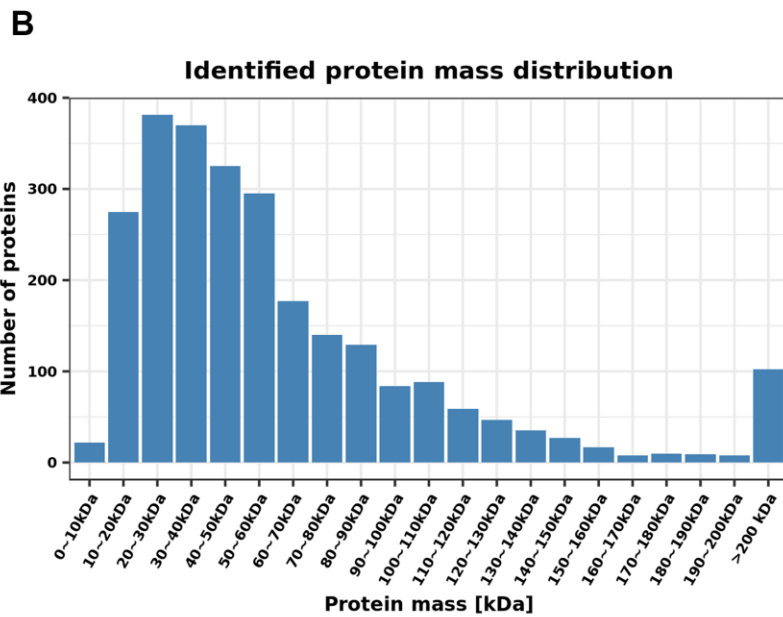
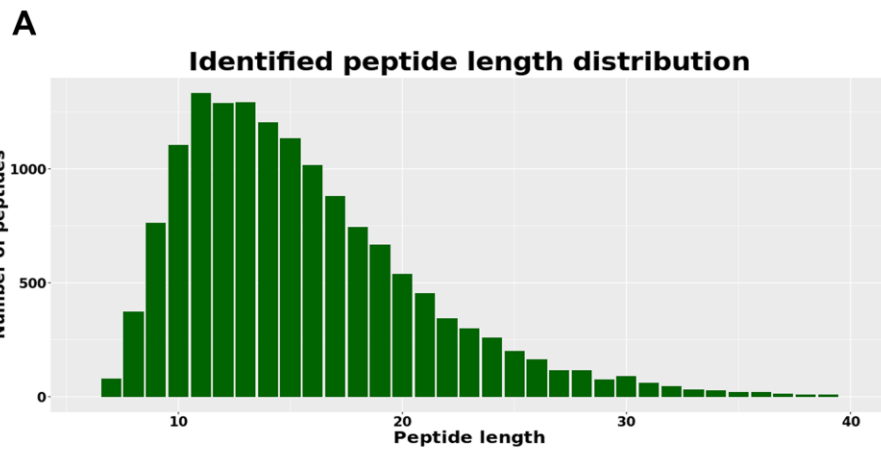
Supplementary Figure 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment clustering analyses of differentially expressed proteins (DEPs) in four quantiles (Q1-Q4).



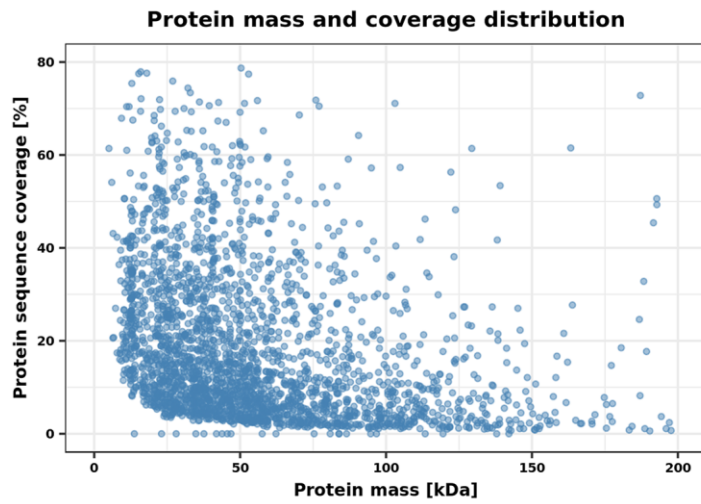
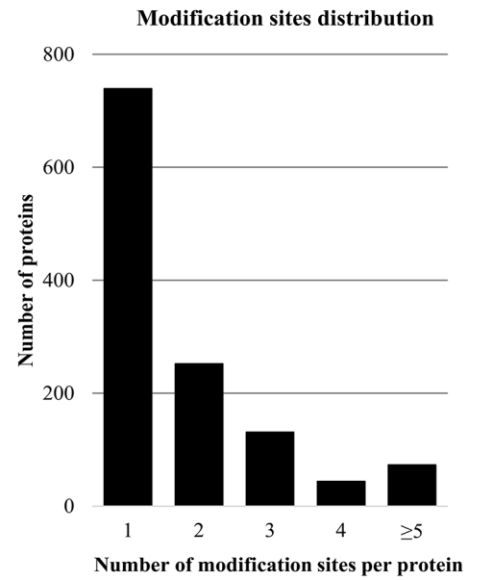
Supplementary Figure 4. Functional enrichment clustering analyses of differentially expressed proteins (DEPs) in four quantiles (Q1-Q4) based on molecular function.



Supplementary Figure 5. Functional enrichment clustering analyses of differentially expressed proteins (DEPs) in four quantiles (Q1-Q4) based on protein domain.



Supplementary Figure 6. (A) The length distribution of peptides identified by mass spectrometry. (B) Molecular weight distribution of all identified proteins. (C) Protein sequence coverage distribution.

A**B**

Supplementary Figure 7. (A) The relationship between the molecular weight of the protein and the coverage determined by mass spectrometry. (B) The number of modification sites of each corresponding protein.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

Supplementary Table 1. All identified and quantified proteins together with quantitative information.

Supplementary Table 2. Clinical characteristics of the primary Sjögren's syndrome (pSS) and normal control (NC) subjects.

	pSS Age Group (30-65 years old)	NC Age Group (30-65 years old)
Number of Subjects	8	10
Age, mean \pm SD years old	46 \pm 6.2	43.2 \pm 6.01
Sex, % female	100%	100%
Disease duration, mean \pm SD years	6.2 \pm 2.7	N/A
% UWSF (\leq 1.5 ml/15 min)	60%	0%
Mean UWSF (ml/15 min)	1.67 \pm 1.2	3.9 \pm 1.7
% + Schirmer's test (\leq 5mm/5min)	80%	0%
Mean Schirmer (mm/5min)	5.2 \pm 7.9	25.4 \pm 8.08
% anti-SSA antibodies	100%	0%
% anti-SSB antibodies	80%	0%
Mean RF	200 \pm 80	N/A

Recruitment of pSS subjects was based on strict American European Consensus Group (AECG) criteria. Non-smoking adult pSS female subjects were screened for oral and ocular dryness, positive labial gland biopsy and positive serological tests for SSA and SSB antibodies. Age-matched, non-smoking adult normal control (NC) females were screened for good general and oral health. *UWSF=Unstimulated whole saliva flow; RF=Rheumatoid factor; N/A= Not applicable.*

Supplementary Table 3. The materials and reagents required for the sample preparation.

Materials and reagents	Supplier
protease inhibitor	Calbiochem
ultrapure water (H ₂ O)	Fisher Chemical
acetonitrile	Fisher Chemical
formic acid	Fluka
phosphorylase inhibitor	Millipore
trypsin	Promega
iodoacetamide	Sigma
urea	Sigma
dithiothreitol	Sigma
TEAB	Sigma
trifluoroacetic acid	Sigma-Aldrich
BCA kit	Beyotime